SOTOPIA: Interactive Evaluation for Social Intelligence in Language Agents

Year: 2023 | Citations: 214 | Authors: Xuhui Zhou, Hao Zhu, Leena Mathur, Ruohong Zhang, Haofei Yu

---

Abstract

Humans are social beings; we pursue social goals in our daily interactions, which is a crucial aspect of social intelligence. Yet, AI systems' abilities in this realm remain elusive. We present SOTOPIA, an open-ended environment to simulate complex social interactions between artificial agents and evaluate their social intelligence. In our environment, agents role-play and interact under a wide variety of scenarios; they coordinate, collaborate, exchange, and compete with each other to achieve complex social goals. We simulate the role-play interaction between LLM-based agents and humans within this task space and evaluate their performance with a holistic evaluation framework called SOTOPIA-Eval. With SOTOPIA, we find significant differences between these models in terms of their social intelligence, and we identify a subset of SOTOPIA scenarios, SOTOPIA-hard, that is generally challenging for all models. We find that on this subset, GPT-4 achieves a significantly lower goal completion rate than humans and struggles to exhibit social commonsense reasoning and strategic communication skills. These findings demonstrate SOTOPIA's promise as a general platform for research on evaluating and improving social intelligence in artificial agents.